Evaluation of Rural Area Classifications Using Statistical Modeling

Stephan J. Goetz and Yicheol Han
Penn State University, Department of AESE, NERCRD and NARDeP[1]

Prepared for the **Workshop on Rationalizing Rural Area Classifications**
National Academy of Sciences and Economic Research Service,
US Department of Agriculture, Washington, DC,
May 9, 2015

This paper assesses the ERS/USDA rural area classification codes using statistical analysis, focusing on whether the codes remain useful and whether the individual categories could be collapsed into a smaller number. We also explore the question of whether there are now better tools for the classification of counties in terms of their status along the rural-urban continuum. Specifically, we adopt recent advances from network science to take explicit advantage of the fact that counties are part of a commuting network within the urban hierarchy and explore in novel ways the notion of labor market areas and how they may in turn be used to classify counties in terms of rurality. Last, we also consider whether outcome measures other than population growth or poverty should be used to evaluate the codes.

Statistical Performance of Existing Codes

We start by examining the "goodness of fit" of the Rural Urban Continuum Code (RUCC), Urban Influence Code (UIC), Rural Urban Commuting Areas (RUCA) and Frontier and Remote (FAR) Area codes in OLS regressions on population (log),

[1] The authors gratefully acknowledge support from USDA/NIFA under grants no. 2012-70002-19385 (NARDeP) and 2014-51150-22094 (NERCRD).

the ratio of employment to population 16 years and older (emp/pop), and population growth. We choose the latter two variables because they are also used by Dijkstra and Ruiz (2010) to identify remote rural regions. In addition, the variable emp/pop is highly correlated with a poverty rates.[2] While this variable represents an indicator at one point in time (i.e., a level ratio), the population growth measure represents a change over time (i.e., it is more of a dynamic indicator). Each of the RUCC and UIC codes is entered as a separate indicator variable (0, 1), while each of the RUCA and FAR codes is entered as a percentage reflecting the population share residing in the respective code. We compare the adjusted R-square values over three periods: 1990, 2000 and 2010, using the most recent data available at the time of this study. Our goal is to compare not only the individual codes against one another but also to compare them over time.

Table 1 reports the results of our initial estimation; Appendix 1-4 presents statistical detail in terms of the individual regression coefficients. The RUCA code performs best (perhaps not surprisingly) in the log population regression, as well as in selected years of the other two variables. In addition, the UIC performs best in the years 1990 and 2000 for emp/pop and in the period 2000-10 in the case of population growth. Our general conclusion from this analysis is that the RUCA code performs reasonably well on the OECD outcome measures considered here, but the "goodness of fit" may be declining over time (in all three periods for emp/pop and in the last period for population growth).

---

[2] The simple correlation coefficients are -0.6854 (in 1990), -0.7929 (in 2000), and -0.7071 (in 2010).

The fact the fit of these regressions is declining over time may be due to the fact that the nature and geographic distribution of the outcome variables is changing, rather than that the rural classification code is no longer relevant. To examine this possibility, we plot the SAIPE-based poverty rate over the years 1989-2013 over time and by the 2013 RUCC code in Figure 1. The 3-D surface shows that the nation was relatively successful in reducing poverty around the year 2000, with more green and blue shadings, and that there is less blue and more red in the most recent years (2013). The graph underscores not only that the poverty rate has risen and remained high since the Great Depression of 2007/08 but also that it has risen in the most urban area. In fact, the graph confirms that poverty has also shifted spatially to become a suburban phenomenon as well. For this reason, it is not a surprise that the goodness of fit between the RUCC and poverty rates has declined over the period examined (Appendix 5).

Next we expand the outcome measures to determine whether the four codes perform differently over the same time periods (Table 2) in terms of identifying rural areas. Here we find that the RUCA performs better than the other codes for the log of population density, the percent of rural population, and the percent of farm area.[3]

Reclassification of the Codes

In this section we explore whether the existing codes could be collapsed to generate a more simple classification scheme. To determine which of the specific

[3] Note that we are ignoring statistical issues such as censoring of the dependent variable.

codes can be grouped we employ a Tukey post-hoc (Figure 2). Those averages that are not significantly different at the 5% level across the regrouped codes are shaded in gray. The largest metropolitan counties (code 1 in both the RUCC and UIC) are significantly different statistically from all the other codes, in all years. While many of the other codes are not significantly different from each other, a key issue with collapsing the codes is that different socio-economic variables would require a different reclassification. And such a reclassification would also have to change across the time periods considered. For these reasons we suggest that reducing the number of codes is not feasible.

An Alternative Classification Approach Based on Network Analysis

The ERS Rural Area Classification codes have served many useful purposes over time, and our above analysis shows that the more recent evolutions, which include more sophisticated measures of commuting patterns, yield a better goodness of fit. Yet all the codes except FAR still rely on negative definitions of rural in the sense that they defined as the residual category (i.e., *not* metro). They also rely on the concepts of adjacency or primary commuting flow to metro area plus a measure of population size related to metro status.

To circumvent these issues we explore how a novel measure that is based on network principles performs in comparison. We use the 3,151 x 3,151 county matrix of commuting flows for the entire US using Census data from 1990, 2000 and 2010 (ACS 2006-10), and exploit this information in two distinct ways. The first is that of calculating the number of distinct labor market areas to which a county is connected. A distinctive feature of this approach is that we can accommodate overlapping labor areas, which we suggest is an improvement over existing methods in which such areas are mutually exclusive. We argue that a county that is connected to a greater number of (more diverse) Labor Market Areas (LMAs) through commuting flows also enjoys more stability over time due to a portfolio effect. We refer to this as our *diversity* measure. The second approach also uses the commuting matrix and adopts a gravity model to measure a county's access to total earnings via commuting to other counties (i.e., total employment weighted by wages). We refer to this as our *proximity* measure.

For the diversity measure, we use the approach in Goetz and Han (2015) to hierarchically sort counties into LMAs based on similarity between commuting link vectors that share a keystone node (county) in the commuting matrix. An innovation of this approach, following Ahn et al. (2010) is that we first sort county commuting links into overlapping LMAs. This has the effect that while commuting links (between counties) are sorted into mutually exclusive LMAs, the corresponding counties can belong to distinctive but overlapping LMAs.[4]

Figure 3 shows a topological rendering of six LMAs in the Northeast region, where we have included only a set of 75 counties for illustrative purposes. To summarize, we first classify counties according to the similarity of their commuting links; then we calculate a diversity measure that reflects the number of distinct but overlapping LMAs to which a county belongs. In this simplified illustration, DC and

[4] See Goetz and Han (2015) for details. Note that because we use links, the smallest number of counties making up one LMA is three (allowing for two or more links).

Baltimore belong to only one LMA while Philadelphia and Salem are parts of three different labor market areas, yielding a diversity score of 3.[5]

For the proximity measure, we calculate the sum of the destination county payrolls that are actually accessed by commuters; a larger proximity measure means more employment and income earning opportunities. Note that unlike the OECD (Dijkstra and Ruiz, 2010) approach or, for example, that of Partridge et al. (2008) ours uses actual commuting data and proportional flows between counties rather than physical distance, or proximity, which may or may not capture actual commuter flows. Specifically, we calculate our measure as:

$$Prox_i = \Sigma_j[(w_{ij} \text{ x } p_j)/(\theta_{ij})^2]$$

where w_{ij} = commuting rate from county i to j

p_j = total annual payroll (in $1,000) in county j,

θ_{ij} = impedance between county i and j.

The data used in these calculations are from the US Census (commuting) the County Business Patterns (total annual payroll)[6] and the Center for Transportation Analysis, Oakridge National Laboratory (impedance).[7]

As the next step in our exploratory analysis we combine these two dimensions of economic access to employment (and other factors such as shopping) into a single measure which we in turn convert into a rural area classification code

[5] Note that these are based on only the subset of 75 counties.
[6] Available at http://www.census.gov/econ/cbp/index.html
[7] Available at http://cta.ornl.gov/transnet/SkimTree.htm

of 10 categories of approximately the same number of counties. Note that we have made a somewhat arbitrary choice between sorting first on diversity and then on proximity, or vice versa. We choose the former in this preliminary work, recognizing that there are more formal ways of accomplishing this task.

Our resulting Network Based County Code (NBCC) classification system is shown in Table 3, where we separate the top-level county codes into Hubs, Hybrids, Hinterlands and Isolated, along with the detailed sub-codes shown within each of these. We reiterate that this scheme (explicitly) uses neither population numbers nor adjacency status. Perhaps of equal interest to policy makers and analysts concerned with rural areas, we do not define rural counties as the residual left over after other codes have been exhausted by urban codes. In particular, we submit that our rural areas consist of the Hinterland counties with limited access to other labor markets (diversity < 7), as well as the 364 Isolated counties which are remote from employment opportunities (proximity<40, regardless of diversity).

Figure 4 shows the NBCC mapped at the county-level. In order to highlight the differences in classification between our new code and one of the existing codes – the RUCC – we include Figure 5. The NBCC classifies a number of counties as non-metro (i.e., hinterland and isolated, or hybrid) that the RUCC classifies as metropolitan. Alternatively, the RUCC classifies a number of counties as non-metro and non-adjacent to metro areas that the NBCC classifies as either a hybrid or a hub county.

Comparative Evaluation of the NBCC

As a final task we examine how the NBCC performs in terms of goodness of fit for the existing outcome measures, as well as select new measures not included earlier. Figures 6 and 7 are box plots for population and emp/pop by NBCC, respectively. Clearly, there is a consistent pattern of decline in population within the three top-level categories, and the isolated counties on average have the smallest populations. The pattern is less clear in the emp/pop figure, suggesting that in future work alternative ways of categorizing the proximity and diversity measures should be explored.

In Table 4 we report adjusted R-square values for the existing rural-urban continuum measures shown in Table 1 and 2. We again shade those cells in which the NBCC yields higher adjusted R-square values than do the existing ERS codes. In the last table (Table 5), we report adjusted R-square values for additional socio-economic measures of social capital (Rupasingha et al. 2006); mental health status (see also Goetz et al. 2014); poverty and poverty change; economic mobility measures from Chetty et al. (2014) as well as child poverty and child poverty change.

The general conclusion here is that our network-based measure of rural status generally underperforms the existing ones in terms of population, emp/pop and population growth, as well as percent rural population and farm area. One exception is the log of population density. At the same, the NBCC tends to outperform the existing measures on the other socioeconomic variables considered, especially the economic mobility variables from Chetty et al. (2014).

Although the adjusted R-square values are not large, in the case of poverty change (1989-00 and 2000-08, which demarks the Great Recessionary period), and the rank-rank slope as well as child poverty rates (1989) the difference is substantial (0.1100 vs. 0.0777, which is the highest, for the RUCA). Perhaps the single most notable finding from this analysis is that the NBCC performs best for the outcomes measured as changes. While this requires further research, we hypothesize that this is the case because networks transmit economic and other shocks or changes, and this is reflected in these results.

Conclusion

We conclude that the existing ERS measures continue to perform reasonably well, depending on the outcome measure used. As discussed at the conference, if they are retained it would be worth considering alternative population cut-off thresholds, however. For future exploration, we submit that a new measure based on emerging tools from network science could be contemplated. As we have shown here in a relatively crude first attempt, such a measure is competitive with the existing ones. This proposed measure does not depend on population size or adjacency concepts, it does not select "rural" as a residual category, and it may in fact be superior when socioeconomic changes as well as economic mobility are selected as the outcome variables.

References

Ahn, Y.-Y., J.P. Bagrow, S. Lehmann. 2010. Link Community reveals multi-scale complexity in networks. *Nature* 446: 761-764

Chetty, R., N. Hendren, P. Kline, E. Saez. 2014. Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. NBER Working Paper No. 19843

Dijkstra, L., V. Ruiz. 2010. Refinement of the OECD regional typology: Economic Performance of Remote Rural Regions. OECD

Goetz, S.J., M. Davlasheridze, Y. Han. 2014. County-level Determinants of Mental Health, 2002-2008. *Social Indicators Research*

Goetz, S.J. and Y. Han. 2015. Identifying Labor Market Areas based on Link Communities. Presented at the AAEA Annual Meetings, San Francisco, CA, draft manuscript.

Partridge, M.D., D.S. Rickman, K. Ali, M.R. Olfert. 2008. Lost in Space: Population Growth in the American Hinterlands and Small Cities. Journal of Economic Geography 8(6): 727-757.

Rupasingha, A., S.J. Goetz and D. Freshwater. 2006. The Production of Social Capital in US Counties, *J. Socio-Economics*

Table 1 Regression results for the Rural area classification codes

Adj. R2	population, log			emp/pop			population growth		
	1990	2000	2010	1990	2000	2010	1990 -00	2000 -10	2010 -13
RUCC	0.297	0.569	0.571	0.091	0.136	0.093	0.096	0.202	0.112
UIC	0.634	0.528	0.532	0.172	0.148	0.098	0.106	0.209	0.112
RUCA	0.680	0.718	0.715	0.171	0.147	0.114	0.132	0.199	0.169
FAR	0.385	0.404	0.416	0.065	0.049	0.007	0.093	0.138	0.031

The highest adjusted R-square values are highlighted.
Source: Authors' calculations

Table 2 Regression results for selected additional outcome measures

Adj. R2	population density, log			% rural population			% farm area		
	1990	2000	2010	1990	2000	2010	1990	2000	2010
RUCC	0.222	0.517	0.533	0.326	0.527	0.528	0.019	0.068	0.075
UIC	0.553	0.513	0.533	0.570	0.437	0.440	0.051	0.062	0.069
RUCA	0.573	0.638	0.715	0.897	0.926	0.910	0.104	0.102	0.135
FAR	0.476	0.497	0.509	0.180	0.230	0.247	0.054	0.051	0.058

The highest adjusted R-square values are highlighted.
Source: Authors' calculations

Table 3 Definition of Network Based County Code, NBCC

Code	Description	No. (2010)
Hub county, diversity ≥ 12		
1	with large potential earnings, 5000 ≤ proximity	303
2	with middle potential earnings, 1000 ≤ proximity < 5000	395
3	with small potential earnings, 40 ≤ proximity < 1000	293
Hybrid county, 7 ≤ diversity < 12		
4	with large potential earnings, 700 ≤ proximity	278
5	with middle potential earnings, 300 ≤ proximity < 700	296
6	with small potential earnings, 40 ≤ proximity < 300	349
Hinterland county, diversity < 7		
7	with large potential earnings, 300 ≤ proximity	321
8	with middle potential earnings, 100 ≤ proximity < 300	333
9	with small potential earnings, 40 ≤ proximity < 100	219
Isolated county		
10	with small potential earnings, proximity < 40	364

Source: Authors' calculations

Table 4 Regression result of NBCC for conventional measures

Adj. R2	population, log			emp/pop			population growth		
	1990	2000	2010	1990	2000	2010	1990 -00	2000 -10	2010 -13
NBCC	0.701	0.603	0.732	0.196	0.138	0.118	0.081	0.159	0.128
Adj. R2	population density, log			% rural population			% farm area		
	1990	2000	2010	1990	2000	2010	1990	2000	2010
NBCC	0.802	0.639	0.809	0.375	0.310	0.453	0.100	0.084	0.100

The highest adjusted R-square values are highlighted. In these cells the goodness of fit is higher than that of any of the other codes.
Source: Authors' calculations

Table 5 Comparison of existing codes with NBCC in additional socio-economic measures

Adj. R2	social capital		PMHD	poverty			poverty change		
	2000	2010	2010	1989	2000	2010	1989 -00	2000 -08	2008 -13
RUCC	0.171	0.215	0.038	0.073	0.131	0.088	0.028	0.031	0.050
UIC	0.175	0.239	0.037	0.128	0.136	0.089	0.049	0.033	0.051
RUCA	0.167	0.276	0.059	0.107	0.127	0.061	0.044	0.068	0.059
FAR	0.181	0.231	0.031	0.066	0.073	0.016	0.023	0.050	0.054
NBCC	0.219	0.240	0.090	0.142	0.144	0.077	0.064	0.088	0.060

Adj. R2	Rank-Rank Slope	Absolute Upward Mobility	Teenage Birth Rate	child poverty			child poverty change		
	2000	2000	2000	1989	2000	2010	1989 -00	2000 -08	2008 -13
RUCC	0.042	0.113	0.043	0.060	0.139	0.132	0.014	0.016	0.087
UIC	0.051	0.114	0.042	0.099	0.145	0.134	0.017	0.017	0.095
RUCA	0.078	0.161	0.055	0.084	0.133	0.112	0.014	0.053	0.084
FAR	0.069	0.146	0.006	0.053	0.083	0.056	0.007	0.007	0.090
NBCC	0.110	0.173	0.063	0.117	0.148	0.101	0.037	0.055	0.119

PMHD: poor mental health days (see Goetz et al. 2014)
The highest adjusted R-square values are highlighted.
Source: Authors' calculations

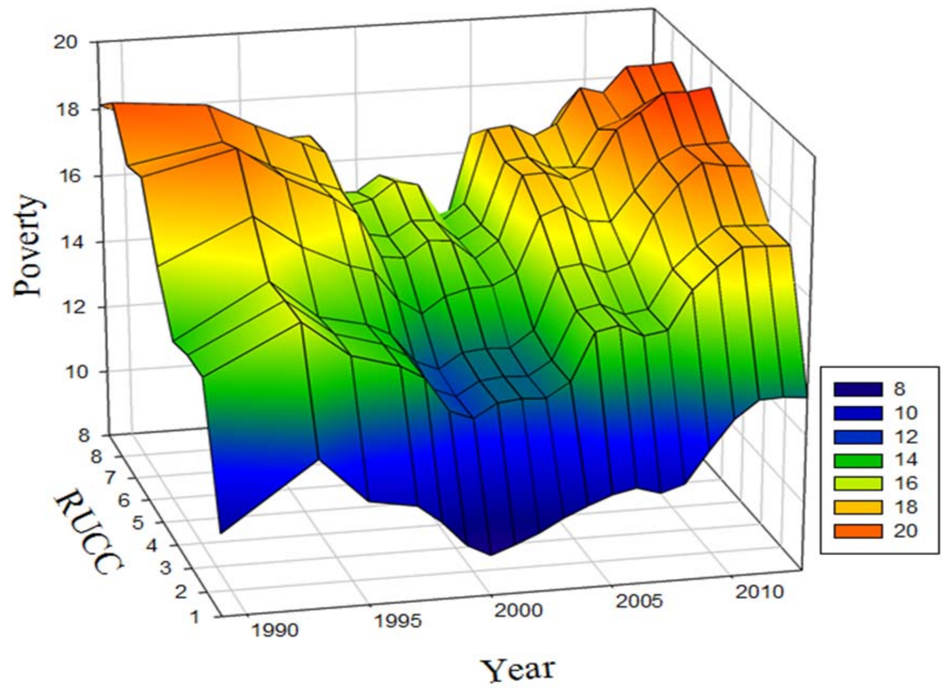

Figure 1 Poverty rate over time (1989-2013) and by RUCC (2013)

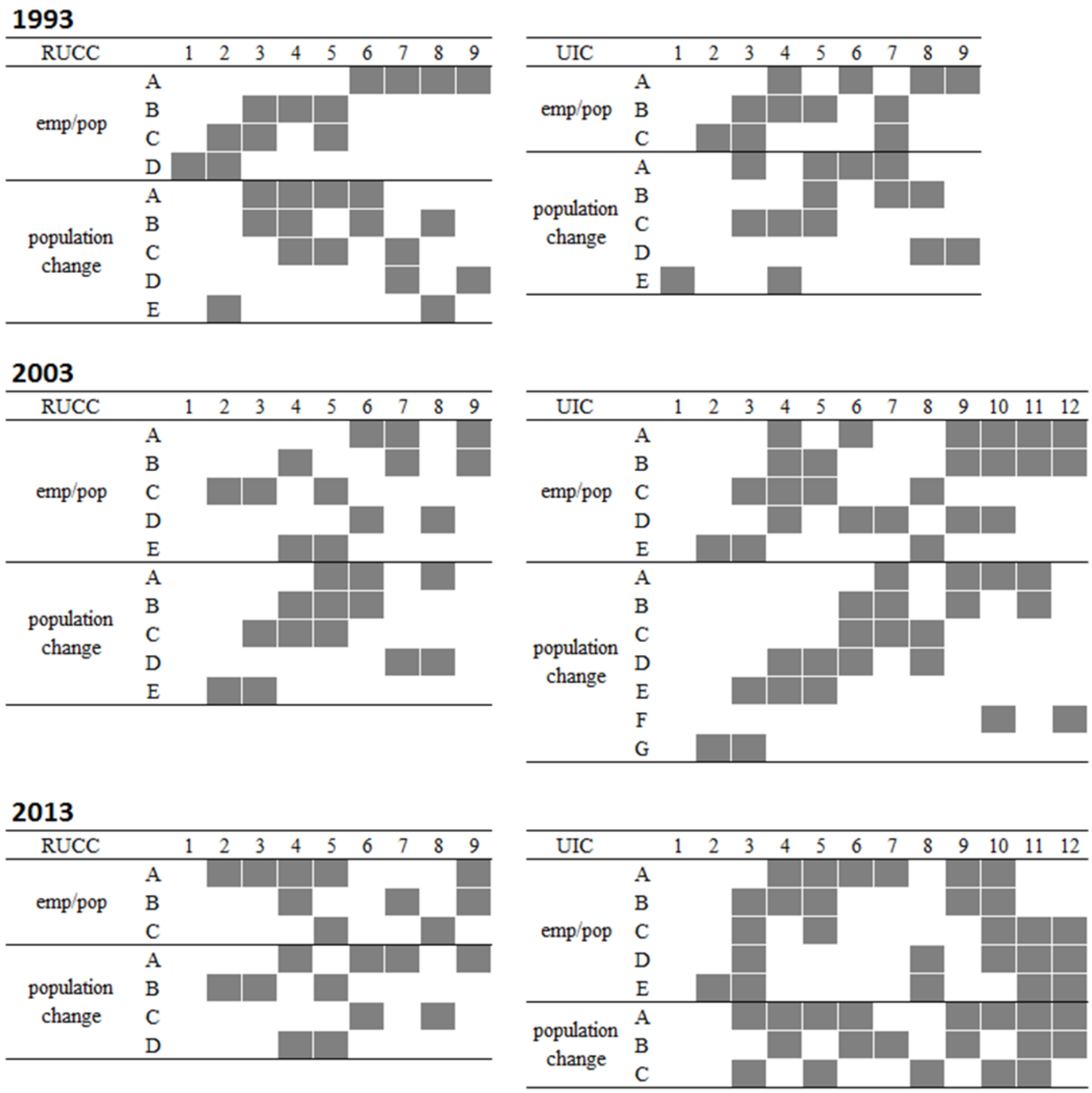

Figure 2 Reclassification of rural-urban codes.

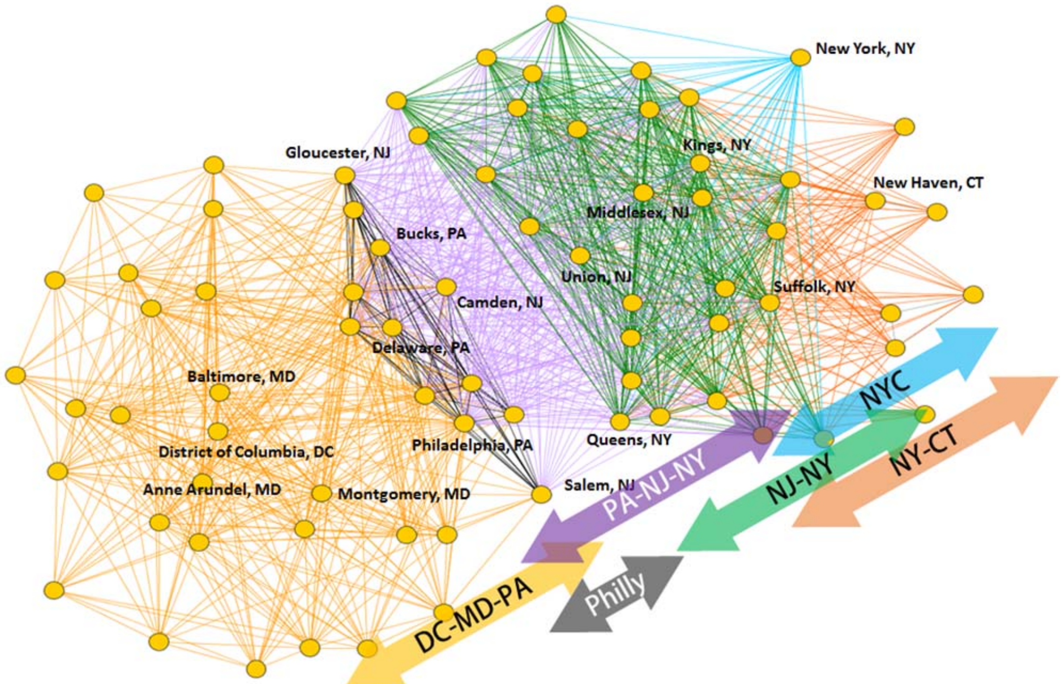

Figure 3 Selected LMAs within the Northeast region's commuting network

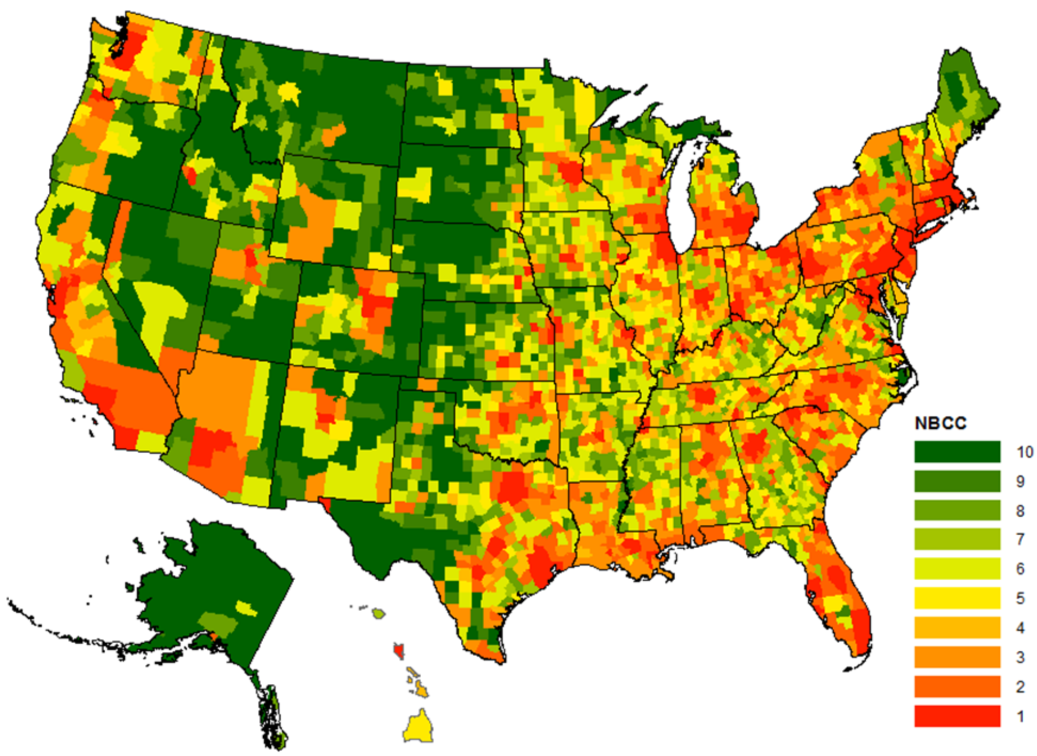

Figure 4 Map of Network Based County Codes, NBCC (2010); see also Table 3

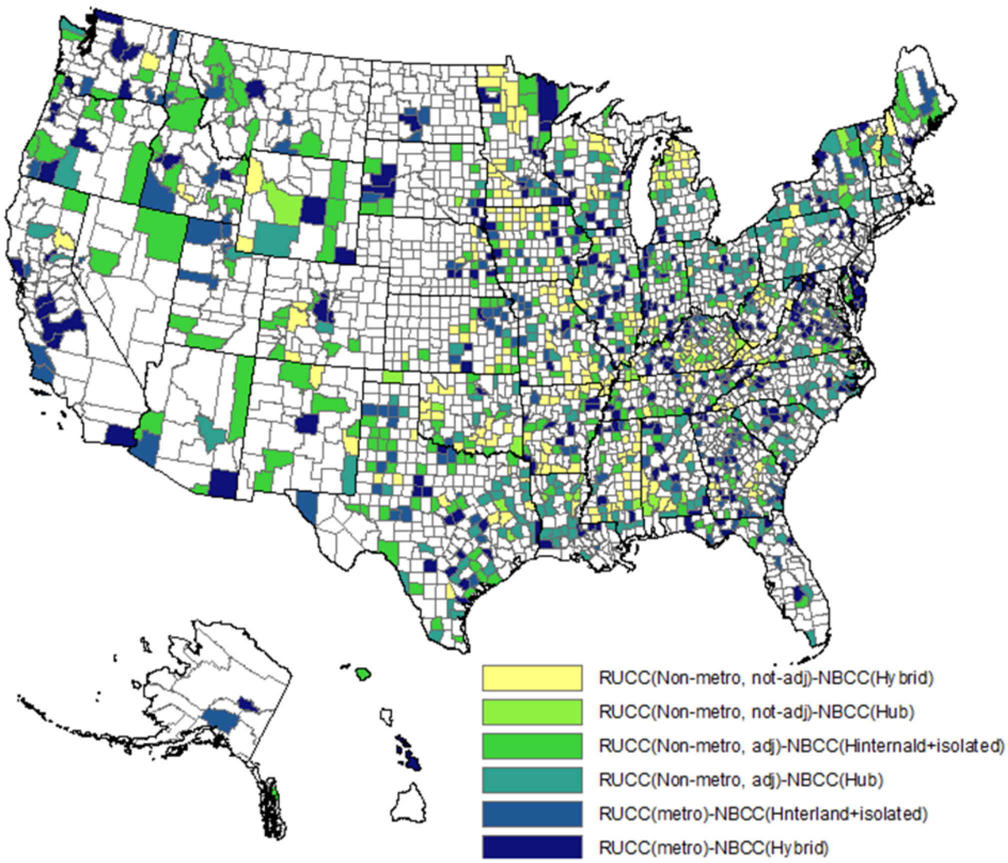

Figure 5 Differences between the NBCC and RUCC

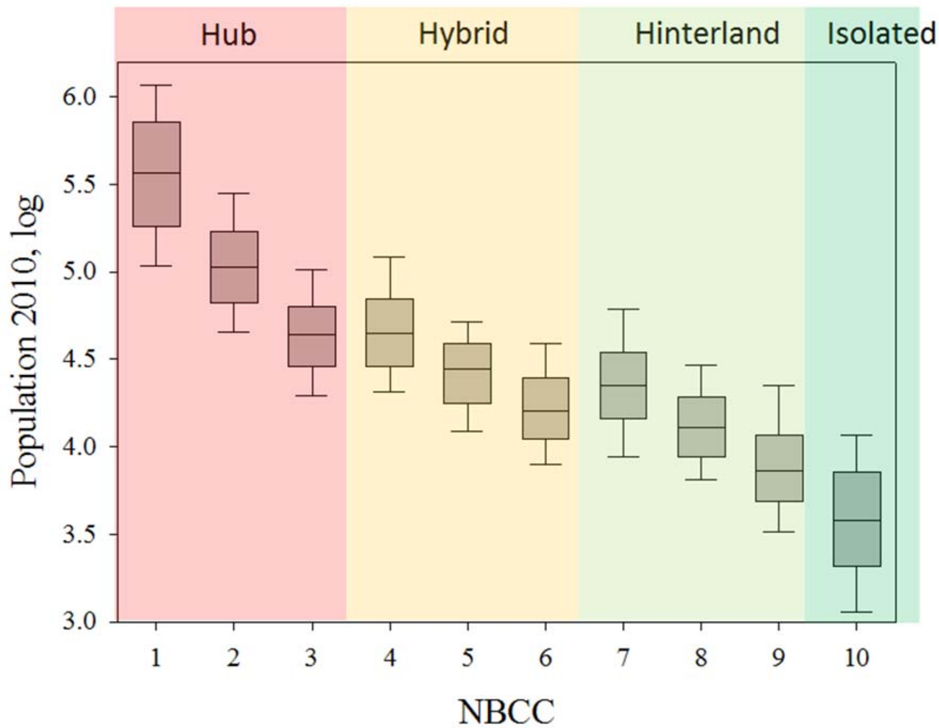

Figure 6 Box plot of population by NBCC

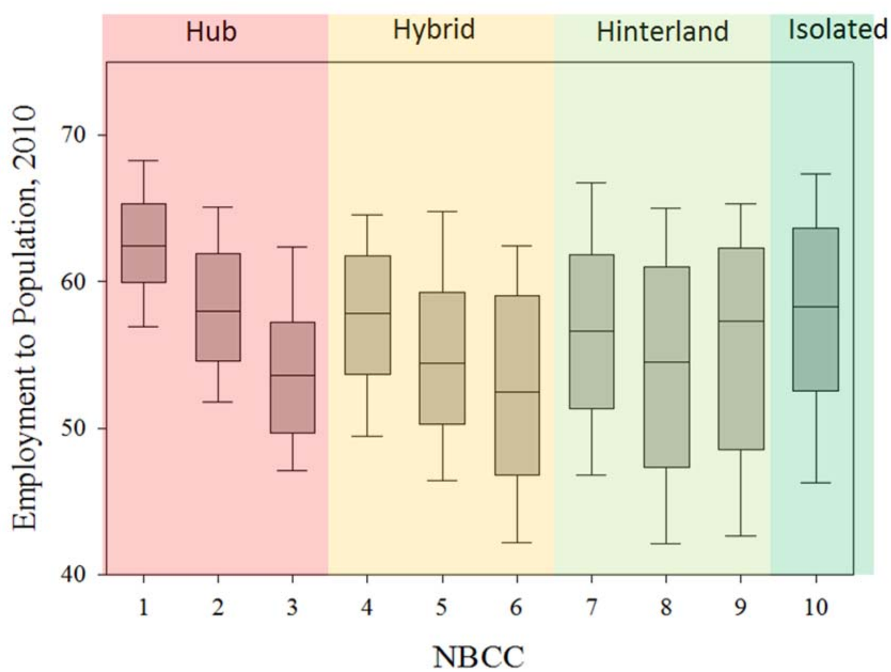

Figure 7 Box plot of Emp/Pop by NBCC

Appendix 1 Regression results using the Rural-Urban Continuum Codes (RUCC)

RUCC	population, log			emp/pop			population growth		
	1990	2000	2010	1990	2000	2010	1990-00	2000-10	2010-13
1	0.19***	0.78***	0.81***	0.16***	0.31***	0.19***	0.27***	0.50***	0.28***
	(12.7)	(37.9)	(40.8)	(9.33)	(14.3)	(8.01)	(9.25)	(18.6)	(9.67)
2	0.43***	0.64***	0.67***	0.15***	0.18***	0.04*	0.17***	0.33***	0.18***
	(19.8)	(36.9)	(38.3)	(7.97)	(8.57)	(1.74)	(8.51)	(16.2)	(6.62)
3	0.31***	0.54***	0.53***	0.10***	0.14***	0.03	0.08***	0.29***	0.13***
	(18.1)	(34.6)	(32.5)	(5.62)	(6.49)	(1.08)	(4.32)	(14.4)	(5.16)
4	0.22***	0.46***	0.45***	0.02	0.04**	-0.03	0.03**	0.18***	0.03*
	(17.3)	(51.2)	(52.3)	(1.14)	(1.98)	(-1.55)	(2.10)	(10.3)	(1.68)
5	0.18***	0.30***	0.28***	0.04**	0.06***	0.04*	0.02	0.12***	0.07***
	(15.5)	(42.6)	(42.2)	(2.24)	(3.19)	(1.77)	(1.46)	(7.67)	(3.93)
6	0.13***	0.43***	0.43***	-0.09***	-0.05**	-0.15***	0.10***	0.20***	-0.03
	(5.49)	(32.0)	(32.9)	(-3.98)	(-2.21)	(-5.27)	(5.14)	(10.2)	(-1.11)
7	0.04*	0.31***	0.31***	-0.09***	0.00	-0.06**	-0.02	0.10***	0.00
	(1.77)	(24.8)	(25.2)	(-3.80)	(-0.08)	(-2.28)	(-0.76)	(5.78)	(0.07)
8	-0.10***	0.10***	0.10***	-0.10***	-0.07***	-0.13***	0.12***	0.10***	-0.08***
	(-6.08)	(8.25)	(8.74)	(-4.74)	(-3.19)	(-5.34)	(5.31)	(5.59)	(-3.64)
_cons	.***	.***	.***	.***	.***	.***	.***	.***	.
	(123)	(200)	(193)	(168)	(150)	(116)	(11.7)	(-6.03)	(-1.56)
Adj. R2	0.2972	0.5694	0.5710	0.0910	0.1359	0.0928	0.0958	0.2021	0.1120

Significance levels: different from zero at *10%, **5%, and ***1% or lower.
Table shows robust standardized coefficients and t-statistics in parentheses.
Source: Authors' calculations

Appendix 2 Regression results with Urban Influence Codes (UIC)

	population, log			emp/pop			population growth		
UIC	1990	2000	2010	1990	2000	2010	1990 -00	2000 -10	2010 -13
1	0.75***	0.81***	0.86***	0.38***	0.33***	0.17***	0.33***	0.54***	0.30***
	(39.9)	(34.5)	(36.9)	(18.3)	(10.5)	(4.83)	(12.4)	(18.2)	(8.55)
2	0.79***	0.82***	0.86***	0.30***	0.23***	0.01	0.25***	0.46***	0.24***
	(52.7)	(36.1)	(35.9)	(13.6)	(6.22)	(0.35)	(11.1)	(16.9)	(5.80)
3	0.24***	0.30***	0.36***	0.06***	0.06***	-0.03	0.08***	0.17***	0.04
	(29.8)	(27.9)	(32.1)	(3.53)	(2.88)	(-1.38)	(5.18)	(9.38)	(1.58)
4	0.17***	0.19***	0.24***	0.04**	-0.01	-0.11***	0.17***	0.16***	-0.01
	(18.1)	(15.9)	(19.5)	(2.39)	(-0.50)	(-4.29)	(7.13)	(8.99)	(-0.66)
5	0.38***	0.49***	0.44***	0.08***	0.04	-0.08***	0.07***	0.21***	0.04*
	(41.6)	(31.9)	(30.8)	(4.45)	(1.43)	(-2.71)	(4.53)	(10.6)	(1.66)
6	0.33***	0.34***	0.35***	0.01	-0.04	-0.16***	0.18***	0.17***	-0.02
	(24.5)	(21.5)	(22.0)	(0.29)	(-1.41)	(-4.75)	(8.27)	(7.93)	(-0.59)
7	0.38***	0.11***	0.12***	0.12***	-0.09***	-0.12***	0.06***	0.09***	-0.06**
	(37.0)	(7.70)	(8.41)	(6.11)	(-3.42)	(-4.46)	(3.81)	(5.15)	(-2.34)
8	0.31***	0.36***	0.37***	0.02	0.10***	0.00	0.06***	0.18***	0.09***
	(25.6)	(20.3)	(21.5)	(0.81)	(3.35)	(0.14)	(2.88)	(8.58)	(2.76)
9		0.22***	0.22***		-0.02	-0.10***		0.08***	0.00
		(16.9)	(17.1)		(-0.68)	(-3.56)		(4.77)	(-0.09)
10		0.04**	0.07***		-0.02	-0.06**		0.03	0.03
		(2.37)	(4.70)		(-0.59)	(-2.00)		(1.32)	(0.76)
11		0.18***	0.17***		0.02	0.00		0.08***	0.01
		(14.6)	(14.7)		(0.95)	(-0.17)		(5.25)	(0.51)
_cons	.***	.***	.***	.***	.***	.***	.***	.***	.**
	(234)	(129)	(119)	(152)	(88.1)	(75.7)	(5.98)	(-6.55)	(-2.00)
Adj. R2	0.6339	0.5280	0.5321	0.1724	0.1479	0.0978	0.1063	0.2087	0.1124

Significance levels: different from zero at *10%, **5%, and ***1% or lower.
Table shows robust standardized coefficients and t-statistics in parentheses.
Source: Authors' calculations

Appendix 3 Regression results with Rural-Urban Commuting Area Codes (RUCA)

RUCA	population, log			emp/pop			population growth		
	1990	2000	2010	1990	2000	2010	1990-00	2000-10	2010-13
1	0.86***	0.92***	0.99***	0.34***	0.29***	0.15***	0.09***	0.36***	0.40***
	(58.7)	(64.7)	(69.7)	(18.9)	(15.6)	(7.20)	(4.91)	(17.0)	(16.2)
2	0.20***	0.20***	0.25***	0.23***	0.20***	-0.09***	0.35***	0.32***	0.02
	(16.3)	(17.8)	(21.6)	(14.5)	(12.4)	(-4.22)	(14.1)	(15.8)	(1.11)
3	0.05***	0.09***	0.10***	0.02	0.02	-0.09***	-0.01	0.06***	-0.04***
	(4.12)	(8.63)	(9.76)	(1.12)	(1.56)	(-4.54)	(-0.82)	(3.54)	(-3.34)
4	0.35***	0.38***	0.43***	0.13***	0.09***	0.05**	0.03	0.16***	0.15***
	(26.7)	(30.6)	(36.9)	(6.17)	(4.11)	(2.47)	(1.59)	(8.46)	(6.79)
5	0.15***	0.15***	0.13***	0.05**	0.03	-0.15***	0.08***	0.06***	-0.04**
	(9.01)	(12.5)	(10.7)	(2.35)	(1.51)	(-7.38)	(3.61)	(3.60)	(-2.38)
6	0.04**	0.08***	0.08***	-0.01	-0.01	-0.09***	0.03**	0.04**	-0.04***
	(2.39)	(6.81)	(6.65)	(-0.52)	(-0.75)	(-6.59)	(2.37)	(1.98)	(-4.22)
7	0.14***	0.15***	0.19***	0.07***	0.02	-0.03	-0.03	0.07***	0.03
	(12.3)	(13.5)	(15.5)	(3.16)	(0.85)	(-1.12)	(-1.36)	(4.24)	(1.26)
8	0.17***	0.13***	0.11***	-0.04**	-0.05***	-0.13***	0.11***	0.03***	-0.02
	(18.4)	(16.9)	(8.62)	(-1.99)	(-2.60)	(-7.29)	(6.26)	(2.62)	(-1.40)
9	0.09***	0.11***	0.09***	-0.02	-0.08***	-0.08***	0.00	0.03	-0.02**
	(9.84)	(12.2)	(6.15)	(-1.25)	(-4.47)	(-5.63)	(0.34)	(1.47)	(-1.99)
_cons	. ***	. ***	. ***	. ***	. ***	. ***	. ***	. ***	. ***
	(297)	(265)	(221)	(196)	(190)	(143)	(11.0)	(-4.88)	(-3.05)
Adj R2	0.6804	0.7175	0.7154	0.1709	0.1465	0.1136	0.1318	0.1992	0.1685

Significance levels: different from zero at *10%, **5%, and ***1% or lower.
Table shows robust standardized coefficients and t-statistics in parentheses.
Source: Authors' calculations

Appendix 4 Regression results with Frontier and Remote Area Codes (FAR)

FAR	population, log			emp/pop			population growth		
	1990	2000	2010	1990	2000	2010	1990-00	2000-10	2010-13
1	-0.27***	-0.28***	-0.29***	-0.20***	-0.16***	-0.11***	-0.22***	-0.27***	-0.15***
	(-10.8)	(-11.4)	(-11.9)	(-5.30)	(-3.96)	(-2.69)	(-7.94)	(-9.16)	(-5.10)
2	0.06**	0.06**	0.07**	0.01	0.04	0.04	0.02	0.06	0.06
	(2.21)	(2.17)	(2.28)	(0.20)	(0.73)	(0.69)	(0.36)	(1.61)	(1.09)
3	-0.45***	-0.45***	-0.45***	-0.07*	-0.11***	-0.02	-0.12***	-0.19***	-0.10*
	(-22.1)	(-22.0)	(-21.9)	(-1.69)	(-2.63)	(-0.38)	(-3.02)	(-6.04)	(-1.82)
_cons	.***	.***	.***	.***	.***	.***	.***	.***	.***
	(376)	(380)	(375)	(354)	(350)	(342)	(39.8)	(29.2)	(11.5)
Adj R2	0.3845	0.4040	0.4157	0.0648	0.0485	0.0074	0.0932	0.1384	0.0312

Significance levels: different from zero at *10%, **5%, and ***1% or lower.
Table shows robust standardized coefficients and t-statistics in parentheses.
Source: Authors' calculations

Appendix 5. Regression results for annual poverty rate by RUCC 2013

RUCC	1989	1993	1995	1997	1998	1999	2000	2001	2002	2003
1	-0.34***	-0.22***	-0.29***	-0.33***	-0.38***	-0.35***	-0.36***	-0.36***	-0.34***	-0.27***
	(-15.2)	(-9.62)	(-12.86)	(-14.47)	(-17.61)	(-16.32)	(-16.92)	(-17.27)	(-15.89)	(-12.41)
2	-0.18***	-0.09***	-0.14***	-0.17***	-0.21***	-0.17***	-0.19***	-0.18***	-0.16***	-0.10***
	(-8.11)	(-3.85)	(-6.39)	(-7.64)	(-9.49)	(-7.56)	(-8.52)	(-8.51)	(-7.55)	(-4.37)
3	-0.15***	-0.06***	-0.11***	-0.13***	-0.16***	-0.13***	-0.14***	-0.14***	-0.13***	-0.07***
	(-7.40)	(-2.70)	(-5.00)	(-5.96)	(-7.65)	(-6.23)	(-6.72)	(-6.73)	(-6.07)	(-3.21)
4	-0.13***	-0.05**	-0.09***	-0.10***	-0.13***	-0.09***	-0.10***	-0.10***	-0.09***	-0.04**
	(-6.83)	(-2.47)	(-4.55)	(-5.29)	(-6.58)	(-4.83)	(-5.42)	(-5.44)	(-4.83)	(-2.20)
5	-0.05***	0.00	-0.03	-0.04**	-0.05**	-0.03	-0.04**	-0.04**	-0.03	0.01
	(-2.78)	(-0.09)	(-1.31)	(-2.13)	(-2.54)	(-1.32)	(-2.10)	(-2.02)	(-1.40)	(0.69)
6	-0.02	0.07***	0.02	-0.01	-0.04	-0.01	-0.02	-0.02	0.00	0.05*
	(-0.85)	(2.75)	(0.72)	(-0.23)	(-1.63)	(-0.47)	(-0.66)	(-0.64)	(-0.14)	(1.90)
7	-0.03	0.05*	0.01	-0.01	-0.03	0.00	-0.01	-0.01	0.00	0.04
	(-1.16)	(1.78)	(0.28)	(-0.47)	(-1.19)	(-0.19)	(-0.48)	(-0.55)	(-0.08)	(1.63)
8	0.02	0.05**	0.03	0.02	0.00	0.01	0.02	0.02	0.02	0.03
	(0.88)	(2.38)	(1.26)	(0.73)	(0.12)	(0.45)	(0.72)	(0.74)	(0.9)	(1.32)
_cons	.***	.***	.***	.***	.***	.***	.***	.***	.***	.***
	(43.1)	(43.6)	(44.5)	(49.2)	(56.0)	(55.0)	(52.5)	(53.8)	(53.9)	(56.7)
Adj R2	0.1152	0.0757	0.0950	0.1097	0.1369	0.1190	0.1297	0.1299	0.1187	0.0903

Significance levels: different from zero at *10%, **5%, and ***1% or lower.
Table shows robust standardized coefficients and t-statistics in parentheses.
Source: Authors' calculations

Appendix 5. Regression results for poverty rate by RUCC 2013 (*continued)*

RUCC	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013
1	-0.25***	-0.31***	-0.32***	-0.31***	-0.30***	-0.29***	-0.24***	-0.21***	-0.21***	-0.20***
	(-11.0)	(-14.2)	(-15.1)	(-14.2)	(-13.7)	(-13.0)	(-10.6)	(-9.54)	(-9.04)	(-9.02)
2	-0.07***	-0.13***	-0.14***	-0.13***	-0.11***	-0.11***	-0.06***	-0.04*	-0.04*	-0.03
	(-3.29)	(-5.79)	(-6.39)	(-5.8)	(-5.23)	(-5.01)	(-2.79)	(-1.96)	(-1.88)	(-1.24)
3	-0.05**	-0.08***	-0.09***	-0.08***	-0.07***	-0.06***	-0.03	-0.01	-0.01	0.00
	(-2.17)	(-3.90)	(-4.18)	(-3.82)	(-3.37)	(-3.02)	(-1.31)	(-0.50)	(-0.42)	(-0.17)
4	-0.02	-0.04**	-0.05**	-0.04**	-0.03	-0.01	0.03	0.04**	0.04*	0.04**
	(-1.22)	(-2.08)	(-2.34)	(-2.08)	(-1.42)	(-0.69)	(1.31)	(2.01)	(1.85)	(2.09)
5	0.02	0.00	0.00	0.00	0.00	-0.01	0.02	0.02	0.03	0.03
	(1.02)	(-0.23)	(-0.03)	(-0.08)	(-0.02)	(-0.27)	(1.03)	(1.27)	(1.31)	(1.51)
6	0.07***	0.03	0.02	0.04	0.06**	0.06**	0.10***	0.12***	0.12***	0.13***
	(2.78)	(1.11)	(0.96)	(1.41)	(2.48)	(2.40)	(3.94)	(4.62)	(4.59)	(5.11)
7	0.06**	0.02	0.02	0.02	0.03	0.03	0.05*	0.06**	0.05**	0.07***
	(2.20)	(0.67)	(0.62)	(0.78)	(1.04)	(1.02)	(1.87)	(2.34)	(2.11)	(2.69)
8	0.05*	0.03	0.03	0.03	0.05**	0.05*	0.05*	0.07***	0.07***	0.08***
	(1.93)	(1.22)	(1.21)	(1.27)	(2.06)	(1.92)	(1.96)	(2.77)	(2.77)	(3.14)
_cons	.***	.***	.***	.***	.***	.***	.***	.***	.***	.***
	(51.7)	(48.3)	(53.1)	(49.3)	(50.7)	(50.2)	(51.5)	(50.5)	(48.1)	(48.8)
Adj R2	0.0861	0.1048	0.1128	0.1070	0.1110	0.1021	0.0875	0.0816	0.0775	0.0806

Significance levels: different from zero at *10%, **5%, and ***1% or lower.
Table shows robust standardized coefficients and t-statistics in parentheses.
Source: Authors' calculations